Workshop of Baltic-Nordic-Ukrainian Network on Survey Statistics 2012

Valmiera, 25 August 2012

INCLUSION PROBABILITIES FOR SUCCESSIVE SAMPLING

Tomas Rudys, Statistics Lithuania

OUTLINE

- Research problem
- Class of order sampling designs
- Successive sampling design
- Expressions for first and second-order inclusion probabilities
- Simulation study results
- Conclusions

Research problem

- Statistics Lithuania recently had a project on "IMPLEMENTATION OF QUALITY IMPROVEMENT ACTIONS FOR THE LABOUR FORCE SURVEY". It was shown that Lithuanian LFS has a successive sampling design.
- On part of the project was to compare inclusion probabilities for successive sampling design and conditional Poisson sampling design.
- The comparison were needed because inclusion probabilities for successive sampling design is difficult to calculate.
- The idea was to analyse the use of inclusion probabilities of conditional Poisson sampling design instead of successive sampling design in the estimation stage.

In this presentation

- The expressions for calculation of first and second-order inclusion probabilities for successive sampling design will be presented.
- The results of simulation study on comparison of inclusion probabilities for both sampling deigns will be shown.

Class of order sampling designs (Rosén, 1996). Let $F_1, F_2, ..., F_N$ be absolutely continuous order distribution functions increasing on $[0, \infty)$. Sampling is carried as follows:

- Independently random variables $Q_1, Q_2, ..., Q_N$ with distributions $F_1, F_2, ..., F_N$ are given.
- Distribution functions $F_1, F_2, ..., F_N$ are associated with the population elements $U = \{1, 2, ..., N\}$ correspondingly.
- $Q_1, Q_2, ..., Q_N$ are realized, and population elements with n smallest Q values constitute the sample.

According to Rosén, successive sampling design is an order sampling design, and its order distribution functions are

$$F_i(t) = 1 - e^{-\theta_i t}, i = 1, 2, ..., N,$$

their density functions are $f_i(t) = F_i'(t) = \theta_i e^{-\theta_i t}$. Here θ_i are given real positive numbers called – intensities. For successive sampling design the intensities are

$$\theta_i = -ln(1 - \lambda_i).$$

Here λ_i is denoted as target inclusion probability.
Simply $\lambda_i, i = 1, ..., N$ are given real numbers which satisfy:
$0 < \lambda_i < 1, \sum_{i=1}^{N} \lambda_i = n.$
It is shown by Rosén that using the order sampling design with fixed distribution shape, inclusion probabilities π_i are approximately equal to given target inclusion probabilities $\lambda_i, i = 1, ..., N$.

Aires (1999) presented expressions for the first and second-order inclusion probabilities of order sampling design. In the case of order sampling design, the probability of element N to belong to the sample is:

$$\pi_N = \int_0^\infty \left(1 - F_n^{N-1}(t)\right) f_N(t)\, dt. \tag{1}$$

Here F_n^N is a distribution function of n-th order statistic of independent random variables $Q_1, Q_2, ..., Q_N$, satisfying the recursive equation

$$F_n^N(t) = F_n^{N-1}(t) + F_N(t)\left(F_{n-1}^{N-1}(t) - F_n^{N-1}(t)\right),$$

and $F_0^N(t) = 1$, for all positive integers N and $t > 0$. Index N may be associated with any population element, and its inclusion probability may be calculated by (1).

Then for the successive sampling design the first-order inclusion probability can be expressed as follows

$$\pi_N = -ln(1-\lambda_N)\int_0^\infty \left(1-F_n^{N-1}(t)\right)(1-\lambda_N)^t dt.$$

Joint inclusion probabilities for any population elements i, j in the n-size order sample are expressed by:

$$\pi_{i,j} = \int_0^\infty \left(1-F_{n-1}^{N-2}(t)\right) f_{max(Q_i,Q_j)}(t)\, dt.$$

In the case of successive sampling, the density function $f_{max(Q_i,Q_j)}(t)$ has a form

$$f_{max(Q_i,Q_j)}(t) = \theta_i e^{-\theta_i t}(1-e^{-\theta_j t}) + (1-e^{-\theta_i t})\theta_j e^{-\theta_j t}.$$

Conditional Poisson sampling design. The elements of the population U are selected independently of each other with probabilities $p_i, i \in U$. The sample size is random, and if it is not equal to n, then the sample is rejected. Sample selection is repeated until the n-size sample is obtained.
The inclusion probability π_i of any element $i \in U, i = 1, ..., N$, can be written as (Aires, 1999):

$$\pi_i = \frac{p_i S_{n-1}^{N-1}(p_1, ..., p_{i-1}, p_{i+1}, ..., p_N)}{S_n^N(p_1, ..., p_N)}.$$

The quantities

$$S_n^N(p_1, ..., p_N) = \sum_{s \in A_n(N)} \prod_{i \in s} p_i \prod_{j \notin s} (1 - p_j)$$

with $N = 0, 1, 2, ...,\ n = 0, ..., N$, and $A_n(N)$-a set of all samples of size n may be calculated recursively by

$$S_n^N(p_1,...,p_N) = p_N S_{n-1}^{N-1}(p_1,...,p_{N-1}) + (1-p_N) S_n^{N-1}(p_1,...,p_{N-1})$$

for $n = 1,...,N-1$ with $S_0^N = (1-p_1)(1-p_2)...(1-p_N)$ and $S_N^N = p_1 p_2 ... p_N$.

The second-order inclusion probability of units i, j to be included in the conditional Poisson sample s, $i \neq j$ is:

$$\pi_{i,j} = \frac{p_i p_j S_{n-2}^{N-2}(p_1,...,p_{i-1},p_{i+1},...,p_{j-1},p_{j+1},...,p_N)}{S_n^N(p_1,...,p_N)}.$$

Simulation results

Let us suppose $N = 5$ size population with element target inclusion probabilities $\lambda = (\lambda_1, \lambda_2, ..., \lambda_N) = (0.1, 0.2, 0.3, 0.5, 0.9)$ is available. We calculate estimates $\hat{\pi}_i$ of the first-order inclusion probabilities π_i for the sample size $n = 2$ elements for both sampling designs. Notice that $\sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \pi_i = 2$. Results are shown in the table.

Table: First-order inclusion probabilities for successive and conditional Poisson sampling designs

i	λ	$\hat{\pi}$	
		Successive	Conditional Poisson
1	0.1	0.087999779247	0.069470260223
2	0.2	0.184249333826	0.154275092936
3	0.3	0.290362518450	0.259990706319
4	0.5	0.540796307599	0.573187732342
5	0.9	0.896591938179	0.943076208178
sum:	2.0	1.999999877303	2.000000000000

Simulation results

We also compute estimates $\hat{\pi}_{i,j}$ of the second order inclusion probabilities $\pi_{i,j}$ for a successive sampling design.

The target inclusion probability vector $\lambda = (0.1, 0.2, 0.3, 0.5, 0.9)$ is given (the same as for first-order inclusion probabilities).

Population size $N = 5$ and sample size $n = 2$ elements.

The control sum is $\sum_{i,j \in U, i<j} \pi_{i,j} = n(n-1)/2 = 1$.

Simulation results

Second-order inclusion probabilities for successive sampling design

i	j				
	1	2	3	4	5
1		0.0036335	0.0059265	0.0121894	0.0662504
2			0.0127577	0.0262163	0.1416418
3				0.0426846	0.2289937
4					0.4597060

Second-order inclusion probabilities for conditional Poisson sampling design

i	j				
	1	2	3	4	5
1		0.0016264	0.0027881	0.0065056	0.0585502
2			0.0062732	0.0146375	0.1317379
3				0.0250929	0.2258364
4					0.5269517

Conclusions

- Simulation results show that inclusion probabilities for both sampling designs are close, but they do not coincide.
- The differences are explained by actual differences of those sampling designs and by approximate numerical integration used to calculate first and second-order inclusion probabilities.

Remark.

Because of recursive expressions under the integrals, the calculation of inclusion probabilities require long computer execution time.

Thank You.